Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Engineering 29 (2012) 986 – 992

Procedia Engineering

www.elsevier.com/locate/procedia

2012 International Workshop on Information and Electronics Engineering (IWIEE)

Chinese Sign Language Synthesis System on Mobile Device

Dan Song*,Xuan Wang,Xinxin Xu

Intelligent Computation Research Center, Harbin Institute of Technology, Shenzhen Graduate School, Shenzhen, 518055, China

Abstract

In this paper, we proposed an effective approach for Chinese Sign Language (CLS) Synthesis System on mobile device. It can translate messages or other text information into Chinese sign languages in real time. In this system, it mainly includes two parts: construction of motion database and sign language synthesis based on text-driven. Construction of motion database is a prerequisite for the Chinese sign language synthesis system. Its motion quality, motion data structure and motion data size together determine the final performance of system. We also introduce a method for motion smoothing in synthesis module. We have integrated two parts on a mobile device, which can render a series of sign language motions according to specific input text from Mobile Internet.

Keywords: Chinese Sign Language Synthesis; mobile device; motion capture; motion editor; VRML

1. Introduction

Chinese sign language is a visual language which expresses certain specific semantics by series of hands and arms motion. It's the most important way to communicate between the deaf and widely used in some specific scenes for privacy or other situation. In our society, there are a large amount of people suffering hearing disability. According to the statistics of relative survey, there are about 21 million deaf in China, of which at the 1-14 age group as high as 1.17 million [1] and the number is increasing rapidly. There is no problem for the deaf to communicate between each other through Chinese sign language but with normal people. To improve the communication barriers, research on sign language synthesis has started back in 1982. Shantz and Poizner proposed their system in [2], and then research on sign language synthesis became a hotspot. A lot of sign language synthesis systems based on PC has been established

* * Corresponding author. *E-mail address*: 373413575@qq.com.

1877-7058 © 2011 Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
doi:10.1016/j.proeng.2012.01.076

for various purposes, for instance, A sign synthesis system named as ViSiCAST [3,4] used motion capture technology to capture the motion data, achieving the conversion from the speech to the British Sign Language and the 3D animation with technology of joint animation and vertex blending. It has been applied to the post office and other public places now. In recent years, with the development of 3D technology and mobile device, sign language expressed as 3D animation on mobile devices is more effective to improve the communication for the deaf [5] in view of mobile device portability. However 3D animation on mobile device has various problems. On one hand the limit of calculation capability and memory results in poor effect which has been achieved perfectly on PC, even some applications cannot run on mobile device without the supporting of hardware. On the other hand, although 3D animation technology is widely used on PC, there are few 3D engines to render motion data on device. The traditional applications mostly adopted browser with VRML plug-in which limits the function extension of application. In this paper we proposed an approach to render 3D animation without browser.

This paper is organized as follow: Section 2 mainly introduces the methods used to construct sign language database such as motion capture technology, static gesture editing and key frames selection strategy etc. Section 3 discusses the implementation of system based on the sign language database. Section 4 is the conclusion.

2. Construction of database

2.1. Avatar modeling

Motion simulation is based on an appropriate model, and modeling is inevitably based on an appropriate definition of avatar body structure which is suitable for inevitably [6].

The definition of avatar adopts hierarchical skeleton structure which can have a true reflection of human motion. At present, the standard of avatar modeling has been more internationally oriented. In system, H-anim standard of VRML [7] is adopted to describe the avatar. The standard defines avatar skeleton as combination of joint points and joints (joint points are connected with joints). According to anatomy it is known that there are more than two hundred degrees-of-freedom, all the degrees-of-freedom determine different human motions. During motion each joint has two forms of movements: rotation and translation. However movement of joint can only be relative to its parent joint rotation and cannot make a shift in space. So it is just defined degree-of-freedom for joint. The parameters of joint which express the static postures of two arms are presented in table 1.

Table 1. Arm joints freedom order

Joints name	Freedom order	Euler angle
Shoulder joints	3	Quaternion
Elbow joints	2	Y,Z
Wrist joints	2	X,Y
Thumb carpometacarpal joints	3	Quaternion
Thumb metacarpophalangeal joints	1	X
Thumb interphalangeal joints	1	X
Remaining four metacarpophalangeal joints	2	Z,X
Remaining four interphalangeal joints	1	Z

There are various definitions of human structures used by researchers. The only difference between systems is the complex of joint hierarchy. This system adopts the common hierarchical skeleton structure [8]. Avatar body is divided into 21 segments (except hands) and the segments are combined with 22 joints. As Chinese sign language is mainly expressed by the motion of arms and hands, the structure of hands should be particularly precise. Facial expressions and movements of body trunk play a secondary role of semantic expression which can make avatar animation more vivid. In system motion characteristics from other parts in addition to hands and arms are neglected which are seen as stationary.

In system, the model of human body is established on the premise of following assumptions:

- Ignore the movements of body trunk and facial expressions, sign language is mainly expressed by upper limb movements
- Human body can be divided into parts, for example, trunk, limbs, legs, and their shape. Each components of human body is rigid, the shape of body trunk and limbs are fixed and cannot be changed
- Each movement of body component are described through the movements in its coordinate system

2.2. Editing of motion

Although a more realistic-looking character animation can be achieved by using motion capture technology, there are also some disadvantages. First, the price of device acquired by motion capture technology is such a high equipment cost which exceeds the capacity of most common scientific research institutions. Second, motion capture asks for high demands on the environment. While processing of capture, it is sensitivity to interference of physical factors. The data for a specific motion should be collected times to times to reach the desired effect. Finally, the reusability of motion data from motion capture technology is poor. For two same actions with nuances should be captured respectively which causes a great deal of data redundancy.

To compensate for the defect of motion capture technology, system adopts motion editor and synthesis technology to improve the reusability for captured motion data. Based on a series of motion data, motion editor mainly make the appropriate adjustment, transformation and mixture operations to remove action bias and redundancy, get the final available motion data which meets realistic standard.

2.2.1. Sign language synthesis

In sign language synthesis module, the specified input text is divided into series of words through a simple word segmentation algorithm. As the sequences of the Chinese sign language is as the same as natural language, the word fragments can be translated into corresponding sign language word animation units and joined together according to the sequence of original input text, and then generate a rich and

powerful motion data. It is flexible to make more rich types of motion data for this method and it is helpful to improve the reusability of unit data.

Motion synthesis is mainly divided into two types. For the connection of two movements with the same type, the same amount of key frame data extracted from the motion data end to end separately can be integrated to form a new smooth motion; For the connection of two different motion types, it is necessary to adopt an effective interpolation algorithm to generate series of consecutive frames between two frames which play a role of smooth for motions. For the smooth connection of human movement data, the human spatial displacement and rotation of each joint should be considered in kinematics.

2.2.1.1. Key frame selection strategy

The size of data acquired by motion capture device is quite large which is not suitable for mobile storage. In order to reduce the size of motion data, an optimization strategy is adopted to cut down frame sequences based on the relative changes between frames.

A sign movement is expressed by a series of continuous frames. To realize the compression of motion data, some relative unimportant frames should be cut down. The strategy is described as follows:

- The sign language consists of some specific gestures which play an important role in sign language recognizing. Such frames should not be abandoned.
- First frame and last frame of two associated gestures are significant to the path of movement. They cannot be discarded.
- If the change of frame which is relative to its two adjacent frames is not so big, the frame can be discarded.

Based on the strategy above, we can calculate the frame weight according to the change relative to its two adjacent frames. To ensure that the first frame and last frame cannot be lost in optimization process, two virtual frames are added which have the same value of first frame and last frame.

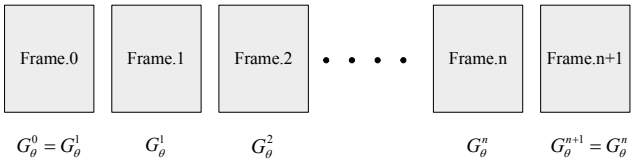

Fig. 1. Expanded sentence frame sequence

Define $\delta_{(i-1,i)}$ as the change between frame i and frame i+1. The formula is defined as shown in equation.

$$\delta_{(i-1,i)} = \sum_{j=1}^{38} (G_{\theta_j}^i - G_{\theta_j}^{i-1})^2 \quad (1)$$

The calculation of frame weight can be expressed by the formula.

$$Q_i = \delta_{(i-1,i)} + \delta_{(i,i+1)} = \sum_{j=1}^{38} (G_{\theta_j}^i - G_{\theta_j}^{i-1})^2 + \sum_{j=1}^{38} (G_{\theta_j}^{i+1} - G_{\theta_j}^i)^2 \quad (2)$$

According the value, some unimportant frames can be discarded to reduce data size.

2.2.2. Rotation interpolation

The changes in human motion posture mainly depend on the rotation movements of each joint. The data captured by the motion capture device is described as Euler angle to denote the change of joints.

However it is difficult to calculate Euler angle interpolation in motion data, and the smooth effect of interpolation is not optimal which often lead to jitter and wrong path. Therefore the transformation from three Euler angles to quaternion [9] is adopted to complete the human motion gesture interpolation through quaternion spherical linear interpolation.

Rotation described by quaternion was proposed by Hamilton in 1843. Let $s = w \in R$ and $\vec{v} = (x, y, z) \in R^3$, quaternion $q \in S^4 = [s, \vec{v}]$, where s, $\vec{v}$ denote the Scalar part of quaternion q and three-dimensional vector [10].

The formula of Sherp spherical linear interpolation algorithm [11] is defined as follows:

If $q_1 = [w_1, x_1, y_1, z_1]$, $q_2 = [w_2, x_2, y_2, z_2]$ are two unit quaternion, spherical interpolation between them is

$$\mathrm{Slerp}(q_1, q_2; u) = \frac{\sin(1-u)\theta}{\sin\theta} q_1 + \frac{\sin u\theta}{\sin\theta} q_2 \tag{3}$$

Where $\theta = \cos^{-1}(w_1 w_2 + x_1 x_2 + y_1 y_2 + z_1 z_2)$.

2.2.3. Construction of database

Based on motion capture technology and motion editing technology, 5590 words in "Chinese Sign Language" and its sequel have been made into corresponding sign language word animations which make up the sign language database. The single animation data includes the data of groups of key frames and corresponding time information. Each animation is stored in a single data file. The size of each motion data is between 2KB and 16KB which is relative to the amount of key frames. The total size of database is about 5.78MB after compression with ZIP format. A table was established to store the index of each motion data in database and a class was encapsulated to search specified animation when given the term. Overall the database is small and portable.

3. CSL synthesis system implementation

At present, there are few ways to render 3D animation on mobile devices. In traditional implementation, researchers employed browser with VRML plug-in to render 3D animation on windows mobile operating system phone OS. The motion data must follow the standard of H-anim. VRML is a standard file format for representing 3D system, but in the recent years, there was no new version published and the features cannot satisfy the increasing command of users. Now the FreeWRL is under developing managed by John A. Stewart [12] which can be supported by OSX, Linux and Windows. However it is not be fully completed. With the FreeWRL application can be transplanted to ios or android operating systems easily

OpenGL ES as a subset of OpenGL 3D graphics application programming interface designed for embedded devices such as mobile phones, PDAs, and video game consoles is widely used in 3D graphics applications recently [13]. It provides some simple models such like rectangle, square etc. which constitute a complex 3D object. Avatar is a complex model based on hierarchical structure and the control of avatar using OpenGL ES is much more difficult. The movement control of joint is limited to its parent and its movement also has impact on its children's motion. If an efficient handler process cannot be implemented, there will be a large calculation before rendering a new animation on screen which probably exerts a considerable influence on interaction effect. The approach adopted in system is using an open dynamic link library from VRML. With the library we can create new applications without browser. In this way the programmer can control the interactive style flexibly to add some new function extensions. The key frame interpolation and the control of library consist of the core of Sign Language Module.

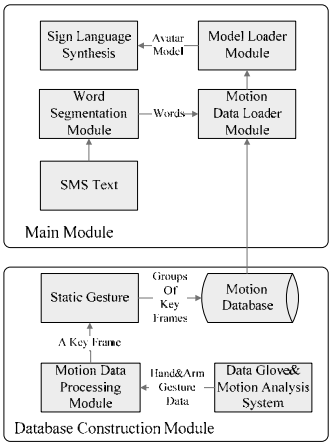

Fig. 2. Mobile sign language synthesis system framework

The Fig.2 represents the framework of Mobile Chinese Sign Language Synthesis System. The system is made up of two parts: main module and database construction module. In database construction module, the technologies mentioned in Section 2 are used to construct the large database. The core of sign language synthesis is completed in main module.

3.1. Main module

In the Main Module, SMS texts are collected in system as input from operating system. Though the word segmentation, SMS are divided into several words which are the inputs of motion data loader module. After loading avatar model, motion data loader searches the corresponding animation datas from motion database and returns the motion datas back to start motion synthesis. The interpolation used has been introduced in Section 2.

3.2. Result of CSL synthesis system

The system runs on HTC HD2 phone with configuration of Qualcomm 8250B 1024MHz CPU, 512M Bytes memory, 4.3 inch screen size, 480×800 resolution and Windows Mobile 6.5 operating system.

Picture (a) in Fig. 4 shows the effect of capturing a message from inbox. While reading the message the avatar model is initializing and waiting for operation from users

Picture (b) to (i) shows the word “围棋” (“Go chess” in English). In the example the content of message is “青岛市围棋联赛战罢第十轮，各队伍纷纷祭出秘密武器”. When user clicks any parts of avatar model on the screen, the model will show Chinese sign language in time. The speed reaches 25 frames per second (fps) on the device mentioned above.

(a) (b) (c) (d) (e) (f) (g)

Fig. 3. Mobile sign language system for SMS

4. Conclusion

The CSL Synthesis System on Mobile Device for the deaf is discussed in this paper. It mainly discussed the technologies and approaches used in system. There are five key techniques: motion capture technology, static editor technology, motion interpolation, key frame selection and animation render approach employed to achieve rendering smooth sign language animation on device. The system is built under some assumptions. For instance, words can be totally expressed by upper limbs movements. Now the system has been achieved on Windows Mobile, joints of avatar are simplified in order to get a faster render effect. Animation can be rendered very smooth and vivid on mobile device. Also there are further improvement of performance considering other new render technology and approach.

References

[1] Jing Wang, Yanfeng Sun, Lichun Wang. Chinese Sign Language Animation System On Mobile Devices. Information Technology and Computer Science (ITCS), 2010 Second International Conference on Digital Object Identifier. 2010.19

[2] Shantz, Poizner. A Computer Program to Synthesize American Sign Language. Behaviour Research Methods and Instrumentation. 1982, 14(5):467~474

[3] Stephen Cox, Michael Lincoln, Judy Tryggvason, et al. Tessa, a system to aid communication with deaf people. Proceedings of the fifth international ACM conference on Assistive technologies, Edinburgh, Scotland. New York, ACM Press, 2002: 205-212.

[4] R. Elliot, J.R.W.Glauert, J.R.Kennaway, I.Marshall. The development of Language processing support for the ViSiCAST projec*t*. Proceedings of the fourth international ACM conference on Assistive technologies, Arlington, Virginia, United States. New York, ACM Press, 2000: 101-108.

[5] HaiMing Huang. Research on motion control of capturing avatar movement[Master Thesis]. Chinese Academy of Sciences, 2007(in chinese)

[6] http://en.wikipedia.org/wiki/VRML

[7] Yan Xu, Jinpeng Huai, Zhaoqi Wang, Data Mining in Hand Gestures and Visual Presentation of the Result, Journal of Computer-Aided Design & Computer Graphics. 2003,15(4):449-453

[8] China Association of the deaf. Chinese Sign Languag. Beijing, HuaXia Press, 1988.

[9] Ken,s. Animating rotateon with quaternion curves. ACM Computer Graphics, 1985,19(3):245~254

[10] Dunn F.,Parberry I. Translated by XueYin Shi. 3D Math Foundation: Development of Graphic and Game. Beijing, Tsinghua University Press, 2005,7:160-172

[11] Yiqiang Chen, Wen Gao, Zhaoqi Wang, Changshui Yang, Dalong Jiang, “Text to Aatar in Multimodal Human Computer Interface”, Asia-Pacific Human Computer Interface (APCHI2002), Vol.2, pp636-643, 2002.

[12] http://en.wikipedia.org/wiki/Opengl_es

[13] http://freewrl.sourceforge.net/